Jover Modrego, J., Fey, N., Harvey, J. N., Lloyd-Jones, G. C., Orpen, A. G., Owen-Smith, G. J. J., Murray, P., Hose, DRJ., Osborne, R., & Purdie, M. (2012). Expansion of the ligand knowledge base for chelating P,P-donor ligands (LKB-PP). *Organometallics*, *31*(15), 5302-5306. https://doi.org/10.1021/om300312t

ORGANOMETALLICS

Article

pubs.acs.org/Organometallics

Expansion of the Ligand Knowledge Base for Chelating P,P-Donor Ligands (LKB-PP)†

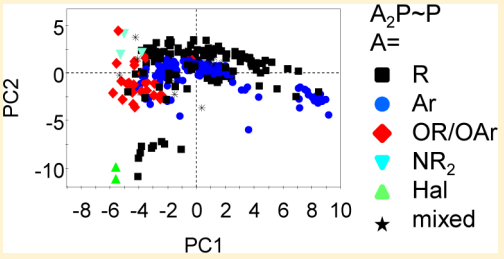
Jesús Jover,[‡,∥] Natalie Fey,*[,‡] Jeremy N. Harvey,[‡] Guy C. Lloyd-Jones,[‡] A. Guy Orpen,[‡] Gareth J. J. Owen-Smith,[‡] Paul Murray,[§,⊥] David R. J. Hose,[§] Robert Osborne,[§] and Mark Purdie[§]

[‡]School of Chemistry, University of Bristol, Cantock's Close, Bristol BS8 1TS, U.K.

[§]AstraZeneca Pharmaceutical Development, Silk Road Business Park, Charter Way, Macclesfield, Cheshire SK10 2NA, U.K.

Supporting Information

ABSTRACT: We have expanded the ligand knowledge base for bidentate P,P- and P,N-donor ligands (LKB-PP, *Organometallics* **2008**, *27*, 1372−1383) by 208 ligands and introduced an additional steric descriptor (nHe_8). This expanded knowledge base now captures information on 334 bidentate ligands and has been processed with principal component analysis (PCA) of the descriptors to produce a detailed map of bidentate ligand space, which better captures ligand variation and has been used for the analysis of ligand properties.



Chelating ligands with phosphorus and nitrogen donor atoms are important in synthetic organometallic chemistry, as they provide a number of features (donor, substituents, backbone) that can be varied in order to modify transition-metal complexes, allowing the fine tuning of complex properties.[1,2] In homogeneous catalysis, such modifications can affect both the rate and selectivity of a reaction, and the quantitative prediction of ligand effects is of particular interest.[2] We have recently reported the development of ligand knowledge bases (LKBs) for a number of different ligand classes, including monodentate ligands with P-[3,4] and C-donor[5] atoms and bidentate ligands with P,P- and P,N-donors.[6] These knowledge bases rely on parameters or descriptors (the terms are used interchangeably here), calculated with a standard DFT approach, which capture the properties of ligands in a range of compounds.

A number of applications for such databases to ligand-driven catalyst optimization can be envisaged, ranging from the illustration of ligand similarities in so-called maps of ligand space[7] to regression models on observed response data.[3,4,6] Comprehensive sampling in screening experiments can be facilitated by using such ligand property maps,[4] derived by principal component analysis (PCA) of the descriptors, and the projection of screening results onto maps can identify "hot spots" of activity due to ligand similarity.[4] More detailed data analysis in terms of descriptors can help to interpret which ligand properties are contributing to observed behavior but can also lead to regression models suitable for making predictions about other ligands.[3,4,6] Training data for such models can be derived from a range of sources, e.g. yield, selectivity, reaction rates,[8] spectroscopic data, and structural features (XRD), as well as computed properties and reactivities.[9]

Our initial LKB-PP[6] accounted for 108 different P,P- and P,N-ligands, capturing some of the trends in ligand electronic and steric features. However, some ligand types were not sampled well or have yet to be accessed experimentally, including, for example, bidentate ligands with electron-withdrawing substituents. Uneven sampling can hamper both ligand selection/test set design and data analysis, making it harder to identify outliers and, if for example the availability, stability, and cost of ligands are of concern, select suitable alternatives. Furthermore, the steric parameter described,[6] He_8_wedge, does not allow for structural changes in response to steric hindrance and is likely to overestimate the size of ligands with flexible backbones.

We have recently demonstrated how substituent and backbone effects on the properties of bidentate P,P-donor ligands can be mapped ($LKB\text{-}PP_{screen}$).[10] This led to the addition of 22 ligands to LKB-PP but also emphasized the need for further ligand variation in the database. Here we report the expansion of LKB-PP by a further 208 ligands, as well as the introduction of a new steric measure (nHe_8), which better captures changes of the ligand bite angle in response to steric hindrance. The ligand descriptors have been processed to produce a map of bidentate ligand space for 334 ligands,[11] evaluating the properties of many ligands for the first time, and this expanded LKB-PP should be useful for ligand screening and subsequent data analysis, as well as prediction of ligand effects.

LIGANDS AND DESCRIPTORS

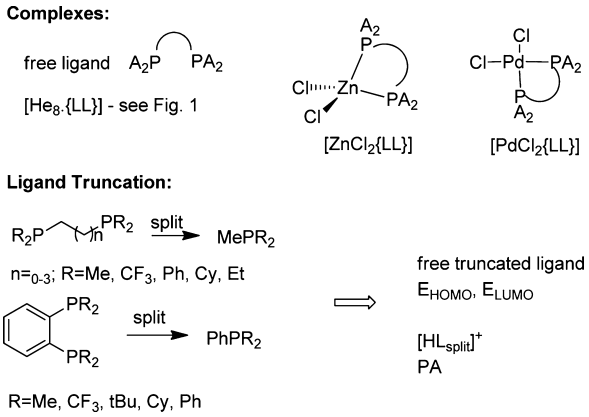
This expansion of the LKB-PP considers a wider range of backbones and substituents, aimed at making the database more relevant to ligand design and experimental screening. Many of the ligands selected have been used and/or

Received: April 17, 2012
Published: July 30, 2012

characterized experimentally, but for subsets of ligands a more systematic evaluation of backbone and substituent effects has also been undertaken, albeit not on the scale of other work reported, which used faster calculations at lower levels of theory.[10,12] Alkyl and aryl substituents are much more common for bidentate P,P-donor ligands than other substituents (OR/OAr, NR_2/NAr_2, Hal), and this continues to be reflected in the ligand set considered here. Details of all ligands, including their unique numbers as used in plots and data tables, have been included in the Supporting Information as Table S1.

The descriptors used have been selected to capture different coordination environments as reported previously.[6] In brief, these have been derived from calculations on the isolated ligand in both free and chelating conformations, the tetrahedral zinc complex [$ZnCl_2${LL}] (LL represents bidentate ligands with P,P- and P,N-donor atoms) and the square-planar palladium(II) complex [$PdCl_2${LL}]. Steric descriptors, discussed in greater detail below, have been determined after optimization of [He_8·{LL}] systems. In addition, ligands have been truncated as described previously[6] to calculate frontier molecular orbital energies and proton affinities for each donor atom (see Table S2 in the Supporting Information for details). Scheme 1 illustrates the complexes considered. A range

Scheme 1. Complexes and Ligand Truncation Used in LKB-PP[a]



[a]See Figure 1 for details of [He_8·{LL}].

of descriptors have been extracted from these calculations, which are detailed in Table S3 of the Supporting Information and may be summarized as follows:

- adduct binding energies for tetrahedral [$ZnCl_2${LL}] and square-planar [$PdCl_2${LL}] complexes
- ligand bite angles in the metal complexes
- structural parameters describing geometry changes of ligands upon complexation and the metal−ligand bond lengths, as well as the geometry of the metal fragments.
- metal fragment charges
- measures of steric bulk (He_8_wedge and nHe_8 steric descriptors)
- frontier molecular orbital energies and ligand proton affinities for each donor atom, derived from truncated ligand fragments (see Table S2 for truncations used)

STERIC DESCRIPTORS

The steric descriptor developed for LKB-PP,[6] He_8_wedge, is calculated as the interaction energy between a ligand in its chelating conformation (from optimization of a zinc complex, [$ZnCl_2${LL}]) and a rigid wedge of eight helium atoms, intended to simulate the steric bulk of other ligands in an octahedral coordination environment. Figure 1 illustrates the

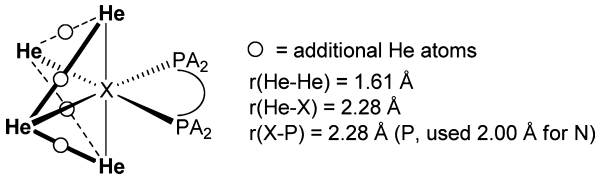


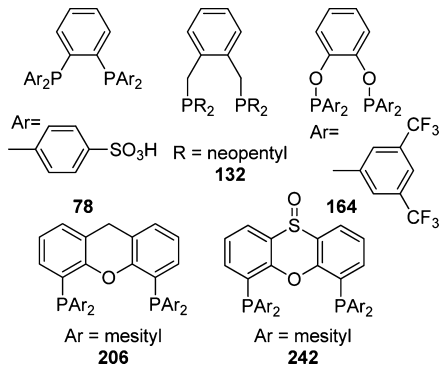
Figure 1. [He_8·{LL}] system used for the calculation of steric descriptors.

[He_8·{LL}] system used, and further details of the calculation setup can be found in the Supporting Information. The He_8_wedge descriptor tends to overestimate the size of some large but "compressible" ligands. This arises because the bite angle[13] adopted in the zinc complex, ∠P−Zn−P, correlates highly[6] with the rather large bite angles adopted in the absence of a coordinating metal, as reported by van Leeuwen et al.[14]

A second steric descriptor, nHe_8, derived from the same starting geometry (Figure 1) as He_8_wedge, has therefore been introduced to LKB-PP to observe the flexible response of ligands to steric hindrance. Instead of completely freezing the positions of donor atoms when calculating the interaction energy between the ligand and the helium wedge, their position is relaxed, subject to the constraint that their distance to the dummy "metal" atom (X) is constant (X−D = 2.28 Å for P and 2.00 Å for N). This allows bite angle changes. While He_8_wedge thus contains steric information about the ligand fixed at a large bite angle, adopted when there is limited steric hindrance on coordination, the nHe_8 descriptor (which is more computationally demanding) allows for relaxation of the backbone and donor atom positions in response to a sterically more crowded environment and this is closer to measuring ligand size in octahedral complexes.

In five cases of large and conformationally flexible ligands (Scheme 2), the nHe_8 optimizations failed to converge and we

Scheme 2. Ligands for Which nHe_8 Descriptor Was Estimated



have instead estimated the nHe_8 parameters from a linear regression model (see the Supporting Information). The relationship between these two steric measures is illustrated in Figure 2.

Ligands have larger values for He_8_wedge than for nHe_8 because fewer constraints are used in the optimization. The

dx.doi.org/10.1021/om300312t | Organometallics 2012, 31, 5302−5306

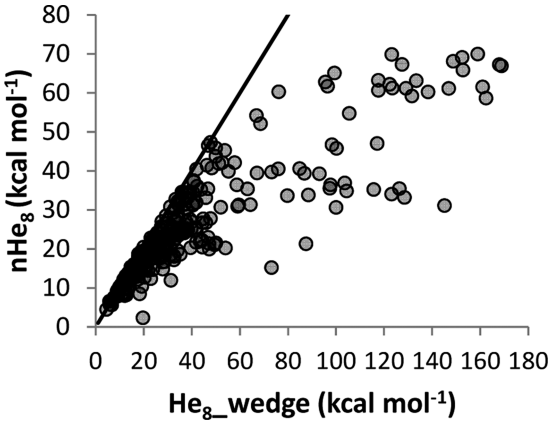


Figure 2. Comparison of steric parameters, He_8_wedge and nHe_8. (The line indicates He_8_wedge = nHe_8.)

difference between them becomes more pronounced for larger ligands (Figure 2). Consideration of both steric parameters can thus better capture the compromise between ligand size and flexibility implicit in different coordination environments.

MAP OF LIGAND SPACE

The most intuitive and accessible display of multiple ligand parameters can be achieved by processing the data to produce a two- or three-dimensional representation of the database, essentially mapping ligand properties. This can be achieved conveniently by processing the data with principal component analysis of the correlation matrix.[15] This statistical approach derives linear combinations of the original descriptors, so-called principal components (PCs), which "distill" the information content of the original database into fewer variables. The detailed composition of PCs can be affected by changes in the ligand or descriptor set, as the approach is sensitive to outliers,[3] but some tentative chemical interpretation of principal components can be achieved where a large number of ligands has been considered,[4] making the expanded LKB-PP more amenable to such analysis.

Figure 3 shows a plot of the first two principal components, which already captures 59% of the information content (variance) in the data set, considering only ligands with P,P-donors. Further PCs can be considered to improve this, and details of PC3 and PC4 are shown in Table 1, while further

Table 1. Descriptor Loadings for PC1−PC4[a]

descriptor	PC1	PC2	PC3	PC4
cumulative contribn, %	39.2	59.4	67.0	74.1
E_{HOMO_P1}	0.163	0.259	−0.328	
E_{HOMO_P2}	0.161	0.265	−0.324	
E_{LUMO_P1}		0.152		0.585
E_{LUMO_P2}		0.148		0.577
PA_{P1}	0.185	0.248	−0.291	
PA_{P2}	0.184	0.255	−0.284	
He_8_wedge	0.261			0.174
nHe_8	0.278			
BE(Zn)		0.318	0.223	
Zn−Cl		0.157		
∠P1−Zn−P2	0.251		0.287	
ΔP1−R(Zn)	0.187		−0.169	
ΔP2−R(Zn)	0.189			
ΔR−P1−R(Zn)	−0.156			
ΔR−P2−R(Zn)	−0.148	0.149		
ΔZn−P1		−0.301	−0.226	0.229
ΔZn−P2		−0.316	−0.184	0.236
Q(Zn)		−0.273	−0.321	
BE(Pd)	−0.185	0.259		
Pd−Cl		0.303		
∠P1−Pd−P2	0.238		0.347	
ΔP1−R(Pd)	0.250			
ΔP2−R(Pd)	0.246			
ΔR−P1−R(Pd)	−0.234			0.191
ΔR−P2−R(Pd)	−0.235		−0.164	0.174
ΔPd−P1	0.278			
ΔPd−P2	0.277			
Q(Pd)	0.195	−0.235	0.206	

[a]Values <|0.145| are not shown; see Table S3 for full data and Table S1 for descriptor definitions (Supporting Information).

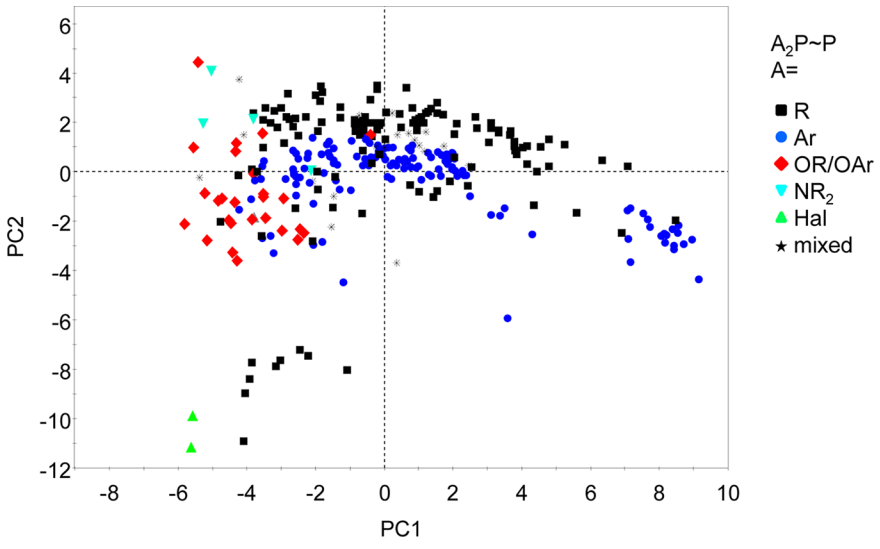


Figure 3. Principal component score plot (PC1 and PC2) for P,P-donor ligands, capturing 59% of the variation in the data set. See the text for a detailed discussion.

dx.doi.org/10.1021/om300312t | Organometallics 2012, 31, 5302−5306

details of the analysis and plots with all ligands identified can be found in the Supporting Information (Table S4, Figure S2). Here we have focused on ligands with two phosphorus(III) donor atoms, but the Supporting Information also includes PCA results for the full set of bidentate ligands, including 24 ligands with P,N-donors (Table S5, Figure S3).

Inspection of the ligand map in Figure 3 suggests that ligands cluster according to their substituents, generating a chemically intuitive representation of changes in ligand properties. Substituent electronic effects are captured mostly by PC2, with halide and perfluorinated alkyl substituents showing negative values of PC2, whereas aryl and alkyl groups give rise to more positive ligand scores on this axis. PC1 shows high loadings for steric descriptors and ligand bite angles, suggesting that backbone and ligand size effects are captured in this direction. This can be illustrated further by color-coding the PC plot according to the number of atoms in the ligand backbone (Figure 4). Short backbones occur mainly at negative PC1,

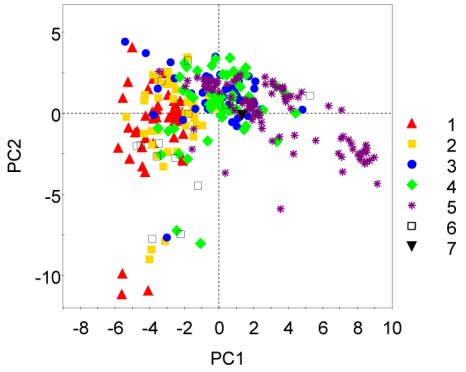


Figure 4. Principal component score plot (PC1 and PC2) for P,P-donor ligands colored according to number of atoms in the backbone.

whereas positive values along this axis are clearly associated with five atoms in the backbone. Longer backbones (six and seven atoms) appear scattered throughout ligand space, which is likely due to their increased flexibility, allowing greater responsiveness to both coordination environment and steric hindrance. Inspection of the descriptor loadings shown in Table 1 can be used to further substantiate these interpretations (see the Supporting Information for details).

This expanded ligand set captures a wider range of ligand properties, and many ligands have been characterized within a coherent data framework for the first time. Nevertheless, there are still areas of ligand space with few examples. Ligands with positive PC1 and negative PC2 scores, i.e. electron-withdrawing substituents and four or five atoms in the backbone, remain undersampled, further supporting the observation that variation in bidentate ligand properties is currently achieved mainly by changes to the ligand backbone and not to substituents. A more complete exploration of different substituents, illustrated in related LKBs,[4,10] is likely to provide access to interesting chemical properties and increase ligand diversity, making such ligands a potential target for novel ligand designs.[7]

SUMMARY AND OUTLOOK

The ligand knowledge base for bidentate P,P- and P,N-donor ligands has been expanded to capture the steric and electronic properties of 334 ligands. In addition, a new steric parameter has been introduced to this LKB-PP, capturing steric effects when the ligand is able to respond to steric hindrance by reducing its bite angle. A map of bidentate ligand space has been derived from principal component analysis of the resulting database. This map illustrates that sampling of ligand space has been improved but also highlights that electron-withdrawing ligands with medium-sized backbones (four to five atoms) are not represented well in comparison with monodentate analogues[4] and so might be interesting targets for novel ligand designs.

While we have demonstrated possible applications of LKBs for data projection and regression models with both an earlier version of this database[6] and for monodentate P-donor ligands (LKB-P),[4] the dearth of suitable experimental data restricts our scope for similar analysis in this case. However, this expanded LKB-PP quantifies the properties of many additional ligands and can be used to select ligands for catalyst screening and optimization, either by visual inspection or using experimental design (DoE) approaches. In addition, ligand effects on the resulting screening data can be interpreted in detail where suitable ligand descriptors are available, and prediction for novel/untested ligands can be attempted from such models. We have recently demonstrated, using proximity on such a ligand map as an indicator of catalytic potential, how novel ligand designs can be assessed and refined with this approach.[7] We are thus optimistic that capturing ligand properties on a quantitative scale can contribute to the design and discovery of improved catalysts, as well as the optimization of industrial processes, both vital steps toward the rational and in silico design of better catalysts.

COMPUTATIONAL DETAILS

Density functional theory calculations were performed in Jaguar[16] and used the standard Becke−Perdew (BP86) density functional.[17] The triple-ζ form of the standard Los Alamos ECP basis set (LACV3P) as implemented in Jaguar was used on the transition-metal atoms, employing the 6-31G* basis for all other atoms. Molecular mechanics (MM) conformational searches used the default MMX force field in PCModel.[18] Principal component analyses were performed in SIMCA-P+.[19]

Details of nondefault program settings have been included in the Supporting Information.

ASSOCIATED CONTENT

Supporting Information

Tables, figures, and a spreadsheet, giving details of ligands, ligand truncations, descriptors, principal component plots identifying all ligands, PC loadings, and ligand scores. This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author

*E-mail: Natalie.Fey@bristol.ac.uk. Tel: (+44) 0117 331 8260. Fax: (+44) 0117 925 1295.

Present Addresses

[‖]ICIQ-Institut Català d'Investigació Química, Avgda. Països Catalans 16, 43007-Tarragona, Spain.

[⊥]CatScI Ltd., CBTC2, Capital Business Park, Wentloog, Cardiff CF3 2PX, U.K.

Notes

The authors declare no competing financial interest.

[†]Development of a Ligand Knowledge Base. 7. See refs 3−6 and 20 for parts 1−6.

■ ACKNOWLEDGMENTS

We thank AstraZeneca Global Pharmaceutical Development for generous financial support of this project. N.F. thanks the EPSRC for the award of an Advanced Research Fellowships (EP/E059376/1). G.C.L.-J. is a Royal Society Wolfson Research Merit Award Holder. We also thank Prof. Paul Pringle, Dr Jason Nichols, and Gemma Bergin for suggesting many additional ligands for consideration.

■ REFERENCES

(1) Birkholz, M. N.; Freixa, Z.; van Leeuwen, P. *Chem. Soc. Rev.* **2009**, *38*, 1099−1118.
(2) Gillespie, J. A.; Dodds, D. L.; Kamer, P. C. J. *Dalton Trans.* **2010**, *39*, 2751−2764.
(3) Fey, N.; Tsipis, A.; Harris, S. E.; Harvey, J. N.; Orpen, A. G.; Mansson, R. A. *Chem. Eur. J.* **2006**, *12*, 291−302.
(4) Jover, J.; Fey, N.; Harvey, J. N.; Lloyd-Jones, G. C.; Orpen, A. G.; Owen-Smith, G. J. J.; Murray, P.; Hose, D. R. J.; Osborne, R.; Purdie, M. *Organometallics* **2010**, *29*, 6245−6258.
(5) Fey, N.; Haddow, M. F.; Harvey, J. N.; McMullin, C. L.; Orpen, A. G. *Dalton Trans.* **2009**, 8183−8196.
(6) Fey, N.; Harvey, J. N.; Lloyd-Jones, G. C.; Murray, P.; Orpen, A. G.; Osborne, R.; Purdie, M. *Organometallics* **2008**, *27*, 1372−1383.
(7) Fey, N.; Garland, M.; Hopewell, J. P.; McMullin, C. L.; Mastroianni, S.; Orpen, A. G.; Pringle, P. G. *Angew. Chem., Int. Ed.* **2012**, *51*, 118−122.
(8) Evans, L. A.; Fey, N.; Harvey, J. N.; Hose, D.; Lloyd-Jones, G. C.; Murray, P.; Orpen, A. G.; Osborne, R.; Owen-Smith, G. J. J.; Purdie, M. *J. Am. Chem. Soc.* **2008**, *130*, 14471−14773.
(9) Fey, N. *Dalton Trans.* **2010**, *39*, 296−310. Fey, N.; Orpen, A. G.; Harvey, J. N. *Coord. Chem. Rev.* **2009**, *253*, 704−722.
(10) Jover, J.; Fey, N. *Dalton Trans.* **2012**, submitted.
(11) Not all descriptors could be calculated for two of the P,N-donor ligands reported previously; therefore, these have been excluded from the present version of LKB-PP.
(12) Burello, E.; Rothenberg, G. *Adv. Synth. Catal.* **2005**, *347*, 1969−1977. Hageman, J. A.; Westerhuis, J. A.; Fruehauf, H.-W.; Rothenberg, G. *Adv. Synth. Catal.* **2006**, *348*, 361−369.
(13) Casey, C. P.; Petrovich, L. M. *J. Am. Chem. Soc.* **1995**, *117*, 6007−6014. Casey, C. P.; Whiteker, G. T.; Melville, M. G.; Petrovich, L. M.; Gavney, J. A.; Powell, D. R. *J. Am. Chem. Soc.* **1992**, *114*, 5535−5543.
(14) Dierkes, P.; van Leeuwen, P. W. N. M. *J. Chem. Soc., Dalton Trans.* **1999**, 1519−1530. van Leeuwen, P. W. N. M.; Kamer, P. C. J.; Reek, J. N. H.; Dierkes, P. *Chem. Rev.* **2000**, *100*, 2741−2770.
(15) Livingstone, D. *A Practical Guide to Scientific Data Analysis*; Wiley: Chichester, U.K., 2009.
(16) *Jaguar 6.0*; Schrödinger LLC, New York, 2005. *Jaguar 7.6*; Schrödinger LLC, New York, 2009.
(17) Becke, A. D. *Phys. Rev. A* **1988**, *38*, 3098−3100. Perdew, J. P. *Phys. Rev. B* **1986**, *33*, 8822−8824. Perdew, J. P. *Phys. Rev. B* **1986**, *34*, 7406. Perdew, J. P.; Zunger, A. *Phys. Rev. B* **1981**, *23*, 5048−5079. Slater, J. C. *Quantum Theory of Molecules and Solids*; McGraw-Hill: New York, 1974; Vol. *4*, The Self-Consistent Field for Molecules and Solids.
(18) Gilbert, K. *PCModel*; Serena Software, Bloomington, IN, 2004.
(19) *SIMCA P+ 12.0.1.0*; Umetrics, 2009.
(20) Fey, N.; Harris, S. E.; Harvey, J. N.; Orpen, A. G. *J. Chem. Inf. Model.* **2006**, *46*, 912−929. Mansson, R. A.; Welsh, A. H.; Fey, N.; Orpen, A. G. *J. Chem. Inf. Model.* **2006**, *46*, 2591−2600.